Paolo Missier

List of Publications by Year in Descending Order

115 papers	1,727 citations	19 h-index	38 g-index
137 ext. papers	2,076 ext. citations	2.2 avg, IF	4.53 L-index

#	Paper	IF	Citations
115	Knowledge-Driven Data Ecosystems Toward Data Transparency. *Journal of Data and Information Quality*, **2022**, 14, 1-12	2.5	4
114	The Right (Provenance) Hammer for the Job: A Comparison of Data Provenance Instrumentation. *Advanced Information and Knowledge Processing*, **2021**, 25-45	0.3	
113	A customisable pipeline for the semi-automated discovery of online activists and social campaigns on Twitter. *World Wide Web*, **2021**, 24, 1-37	2.9	0
112	Using Wearable Activity Trackers to Predict Type 2 Diabetes: Machine Learning-Based Cross-sectional Study of the UK Biobank Accelerometer Cohort. *JMIR Diabetes*, **2021**, 6, e23364	2.7	3
111	Machine learning in predicting respiratory failure in patients with COVID-19 pneumonia-Challenges, strengths, and opportunities in a global health emergency. *PLoS ONE*, **2020**, 15, e0239172	3.7	19
110	Capturing and querying fine-grained provenance of preprocessing pipelines in data science. *Proceedings of the VLDB Endowment*, **2020**, 14, 507-520	3.1	7
109	Abstracting PROV provenance graphs: A validity-preserving approach. *Future Generation Computer Systems*, **2020**, 111, 352-367	7.5	2
108	Why-Diff: Exploiting Provenance to Understand Outcome Differences From Non-Identical Reproduced Workflows. *IEEE Access*, **2019**, 7, 34973-34990	3.5	1
107	Increasing phenotypic annotation improves the diagnostic rate of exome sequencing in a rare neuromuscular disorder. *Human Mutation*, **2019**, 40, 1797-1812	4.7	9
106	Targeted therapies for congenital myasthenic syndromes: systematic review and steps towards a treatabolome. *Emerging Topics in Life Sciences*, **2019**, 3, 19-37	3.5	25
105	Toward trusted open data and services. *Internet Technology Letters*, **2019**, 2, e69	1.3	4
104	Report on the First International Workshop on Incremental Re-computation. *SIGMOD Record*, **2019**, 47, 35-38	1.1	
103	A Customisable Pipeline for Continuously Harvesting Socially-Minded Twitter Users. *Lecture Notes in Computer Science*, **2019**, 91-106	0.9	1
102	Toward a Decentralized, Trust-Less Marketplace for Brokered IoT Data Trading Using Blockchain **2019**,		21
101	VazaDengue: An information system for preventing and combating mosquito-borne diseases with social networks. *Information Systems*, **2018**, 75, 26-42	2.7	17
100	Selective and Recurring Re-computation of Big Data Analytics Tasks: Insights from a Genomics Case Study. *Big Data Research*, **2018**, 13, 76-94	3.7	3
99	Analyzing Social Network Images with Deep Learning Models to Fight Zika Virus. *Lecture Notes in Computer Science*, **2018**, 605-610	0.9	3

98	Provenance Annotation and Analysis to Support Process Re-computation. *Lecture Notes in Computer Science*, **2018**, 3-15	0.9	
97	Leveraging Blockchain to Enable Smart-Health Applications **2018**,		9
96	Versioned-PROV: A PROV Extension to Support Mutable Data Entities. *Lecture Notes in Computer Science*, **2018**, 87-100	0.9	2
95	TAPER: query-aware, partition-enhancement for large, heterogenous graphs. *Distributed and Parallel Databases*, **2017**, 35, 85-115	0.9	5
94	Mind my value **2017**,		18
93	Recruiting from the Network: Discovering Twitter Users Who Can Help Combat Zika Epidemics. *Lecture Notes in Computer Science*, **2017**, 437-445	0.9	10
92	Preserving the Value of Large Scale Data Analytics over Time Through Selective Re-computation. *Lecture Notes in Computer Science*, **2017**, 65-77	0.9	1
91	Tracking Dengue Epidemics Using Twitter Content Classification and Topic Modelling. *Lecture Notes in Computer Science*, **2016**, 80-92	0.9	17
90	Provenance and data differencing for workflow reproducibility analysis. *Concurrency Computation Practice and Experience*, **2016**, 28, 995-1015	1.4	19
89	Scalable and efficient whole-exome data processing using workflows on the cloud. *Future Generation Computer Systems*, **2016**, 65, 153-168	7.5	8
88	Data trajectories: tracking reuse of published data for transitive credit attribution. *International Journal of Digital Curation*, **2016**, 11, 1-16	0.9	4
87	The Lifecycle of Provenance Metadata and Its Associated Challenges and Opportunities. *Springer Proceedings in Business and Economics*, **2016**, 127-137	0.2	3
86	DataONE: A Data Federation with Provenance Support. *Lecture Notes in Computer Science*, **2016**, 230-234	0.9	3
85	Analyzing Provenance Across Heterogeneous Provenance Graphs. *Lecture Notes in Computer Science*, **2016**, 57-70	0.9	5
84	Clustering provenance facilitating provenance exploration through data abstraction **2016**,		5
83	ProvGen: Generating Synthetic PROV Graphs with Predictable Structure. *Lecture Notes in Computer Science*, **2015**, 16-27	0.9	7
82	Access control and view generation for provenance graphs. *Future Generation Computer Systems*, **2015**, 49, 8-27	7.5	7
81	**2015**,		23

80	YesWorkflow: A User-Oriented, Language-Independent Tool for Recovering Workflow Information from Scripts. *International Journal of Digital Curation*, **2015**, 10, 298-313	0.9	47
79	ProvAbs: Model, Policy, and Tooling for Abstracting PROV Graphs. *Lecture Notes in Computer Science*, **2015**, 3-15	0.9	8
78	Provenance Storage, Querying, and Visualization in PBase. *Lecture Notes in Computer Science*, **2015**, 239-241	0.9	1
77	Provenance-Based Searching and Ranking for Scientific Workflows. *Lecture Notes in Computer Science*, **2015**, 209-214	0.9	1
76	SVI: A Simple Single-Nucleotide Human Variant Interpretation Tool for Clinical Use. *Lecture Notes in Computer Science*, **2015**, 180-194	0.9	2
75	A Platform for Analysing Stream and Historic Data with Efficient and Scalable Design Patterns **2014**,		2
74	From Scripted HPC-Based NGS Pipelines to Workflows on the Cloud **2014**,		5
73	Tweet My Street: A Cross-Disciplinary Collaboration for the Analysis of Local Twitter Data. *Future Internet*, **2014**, 6, 378-396	3.3	14
72	Report from the second workshop on scalable workflow enactment engines and technology (SWEET13). *SIGMOD Record*, **2014**, 42, 73-77	1.1	0
71	On strategies for budget-based online annotation in human activity recognition **2014**,		3
70	Measuring the impact of cognitive distractions on driving performance using time series analysis **2014**,		4
69	Distilling structure in Taverna scientific workflows: a refactoring approach. *BMC Bioinformatics*, **2014**, 15 Suppl 1, S12	3.6	6
68	The PBase Scientific Workflow Provenance Repository. *International Journal of Digital Curation*, **2014**, 9, 28-38	0.9	16
67	Measuring Population-Based Completeness for Single Nucleotide Polymorphism (SNP) Databases. *Studies in Computational Intelligence*, **2014**, 173-182	0.8	2
66	Forget Dimensions: Define Your Information Quality Using Quality View Patterns. *Synthese Library*, **2014**, 25-41	0.2	
65	Why linked data is not enough for scientists. *Future Generation Computer Systems*, **2013**, 29, 599-611	7.5	160
64	The W3C PROV family of specifications for modelling provenance metadata **2013**,		73
63	Extracting PROV provenance traces from Wikipedia history pages **2013**,		9

62	Report from the first workshop on scalable workflow enactment engines and technology (SWEETU2). *SIGMOD Record*, **2013**, 41, 60-64	1.1	
61	Measuring Data Completeness for Microbial Genomics Database. *Lecture Notes in Computer Science*, **2013**, 186-195	0.9	4
60	Reference Architectures to Measure Data Completeness across Integrated Databases. *Lecture Notes in Computer Science*, **2013**, 216-225	0.9	2
59	Predicting the Execution Time of Workflow Activities Based on Their Input Features **2012**,		23
58	A PROV Encoding for Provenance Analysis Using Deductive Rules. *Lecture Notes in Computer Science*, **2012**, 67-81	0.9	6
57	Detecting Duplicate Records in Scientific Workflow Results. *Lecture Notes in Computer Science*, **2012**, 126-138	0.9	1
56	Modelling Provenance Using Structured Occurrence Networks. *Lecture Notes in Computer Science*, **2012**, 183-197	0.9	
55	. *IEEE Internet Computing*, **2011**, 15, 40-48	2.4	14
54	Simulating Taverna workflows using stochastic process algebras. *Concurrency Computation Practice and Experience*, **2011**, 23, 1920-1935	1.4	1
53	The Open Provenance Model core specification (v1.1). *Future Generation Computer Systems*, **2011**, 27, 743-756	7.5	396
52	Workflows to open provenance graphs, round-trip. *Future Generation Computer Systems*, **2011**, 27, 812-819	7.5	10
51	Achieving reproducibility by combining provenance with service and workflow versioning **2011**,		14
50	The panel of experts cloud pattern **2011**,		2
49	An experimental workflow development platform for historical document digitisation and analysis **2011**,		5
48	eScience **2011**, 701-736		2
47	Workflows for Information Integration in the Life Sciences. *Lecture Notes in Computer Science*, **2011**, 215-225	0.9	1
46	Linking multiple workflow provenance traces for interoperable collaborative science **2010**,		20
45	Functional Units: Abstractions for Web Service Annotations **2010**,		5

44	ERGOT: A Semantic-Based System for Service Discovery in Distributed Infrastructures **2010**,		14
43	The Evolution of myExperiment **2010**,		5
42	Why Linked Data is Not Enough for Scientists **2010**,		31
41	A Comparison of Using Taverna and BPEL in Building Scientific Workflows: the case of caGrid. *Concurrency Computation Practice and Experience*, **2010**, 22, 1098-1117	1.4	5
40	A formal semantics for the Taverna 2 workflow model. *Journal of Computer and System Sciences*, **2010**, 76, 490-508	1	19
39	Fine-grained and efficient lineage querying of collection-based workflow provenance **2010**,		43
38	Taverna, Reloaded. *Lecture Notes in Computer Science*, **2010**, 471-481	0.9	71
37	Janus: From Workflows to Semantic Provenance and Linked Open Data. *Lecture Notes in Computer Science*, **2010**, 129-141	0.9	31
36	Understanding Collaborative Studies through Interoperable Workflow Provenance. *Lecture Notes in Computer Science*, **2010**, 42-58	0.9	12
35	A Linked Data Approach to Sharing Workflows and Workflow Results. *Lecture Notes in Computer Science*, **2010**, 340-354	0.9	2
34	**2009**,		2
33	Time-completeness trade-offs in record linkage using adaptive query processing **2009**,		6
32	The data playground: An intuitive workflow specification environment. *Future Generation Computer Systems*, **2009**, 25, 453-459	7.5	8
31	Incorporating Domain-Specific Information Quality Constraints into Database Queries. *Journal of Data and Information Quality*, **2009**, 1, 1-31	2.5	6
30	Building Scientific Workflow with Taverna and BPEL: A Comparative Study in caGrid. *Lecture Notes in Computer Science*, **2009**, 118-129	0.9	16
29	Medical image processing workflow support on the EGEE grid with taverna **2009**,		8
28	Data Provenance in Scientific Workflows **2009**, 46-59		1
27	Brokering infrastructure for minimum cost data procurement based on quality-quantity models. *Decision Support Systems*, **2008**, 45, 95-109	5.6	7

26	An ontology-based approach to handling information quality in e-Science. *Concurrency Computation Practice and Experience*, **2008**, 20, 253-264	1.4	9
25	Information quality in proteomics. *Briefings in Bioinformatics*, **2008**, 9, 174-88	13.4	25
24	Exploiting Provenance to Make Sense of Automated Decisions in Scientific Workflows. *Lecture Notes in Computer Science*, **2008**, 174-185	0.9	2
23	Data Lineage Model for Taverna Workflows with Lightweight Annotation Requirements. *Lecture Notes in Computer Science*, **2008**, 17-30	0.9	32
22	Grid metadata management: Requirements and architecture **2007**,		3
21	Managing information quality in e-science **2007**,		9
20	Requirements and Services for Metadata Management. *IEEE Internet Computing*, **2007**, 11, 17-25	2.4	12
19	Taverna Workflows: Syntax and Semantics **2007**,		46
18	Architectural Patterns for the Semantic Grid **2007**, 119-134		
17	Accelerating Disease Gene Identification Through Integrated SNP Data Analysis **2007**, 215-230		1
16	Metadata Management in S-OGSA. *Lecture Notes in Computer Science*, **2007**, 712-719	0.9	
15	Semantic Support for Meta-Scheduling in Grids **2007**, 169-183		4
14	Managing Information Quality in e-Science Using Semantic Web Technology. *Lecture Notes in Computer Science*, **2006**, 472-486	0.9	7
13	Towards the Management of Information Quality in Proteomics **2006**,		2
12	An overview of S-OGSA: A Reference Semantic Grid Architecture. *Web Semantics*, **2006**, 4, 102-115	2.9	50
11	Clustering Web pages based on their structure. *Data and Knowledge Engineering*, **2005**, 54, 279-299	1.5	40
10	Provider issues in quality-constrained data provisioning **2005**,		1
9	Managing Information Quality in e-Science: A Case Study in Proteomics. *Lecture Notes in Computer Science*, **2005**, 423-432	0.9	4

8	Fine-grain web site structure discovery **2003**,		7
7	Improving Data Quality in Practice: A Case Study in the Italian Public Administration. *Distributed and Parallel Databases*, **2003**, 13, 135-160	0.9	13
6	Electronic Government. *Lecture Notes in Computer Science*, **2003**,	0.9	21
5	The Service to Businesses Project: Improving Government-to-Business Relationships in Italy. *Lecture Notes in Computer Science*, **2003**, 468-471	0.9	2
4	CitiTime: a system for rapid creation of portable next-generation telephony services. *Computer Networks*, **2001**, 35, 579-595	5.4	1
3	Java call control, coordination, and transactions **2000**, 38, 108-114		9
2	A framework for analyzing virtual enterprise infrastructure **1999**,		4
1	An Overview of S-OGSA: A Reference Semantic Grid Architecture. *SSRN Electronic Journal*,	1	1